

CHAPTER III

RESEARCH METHODOLOGY

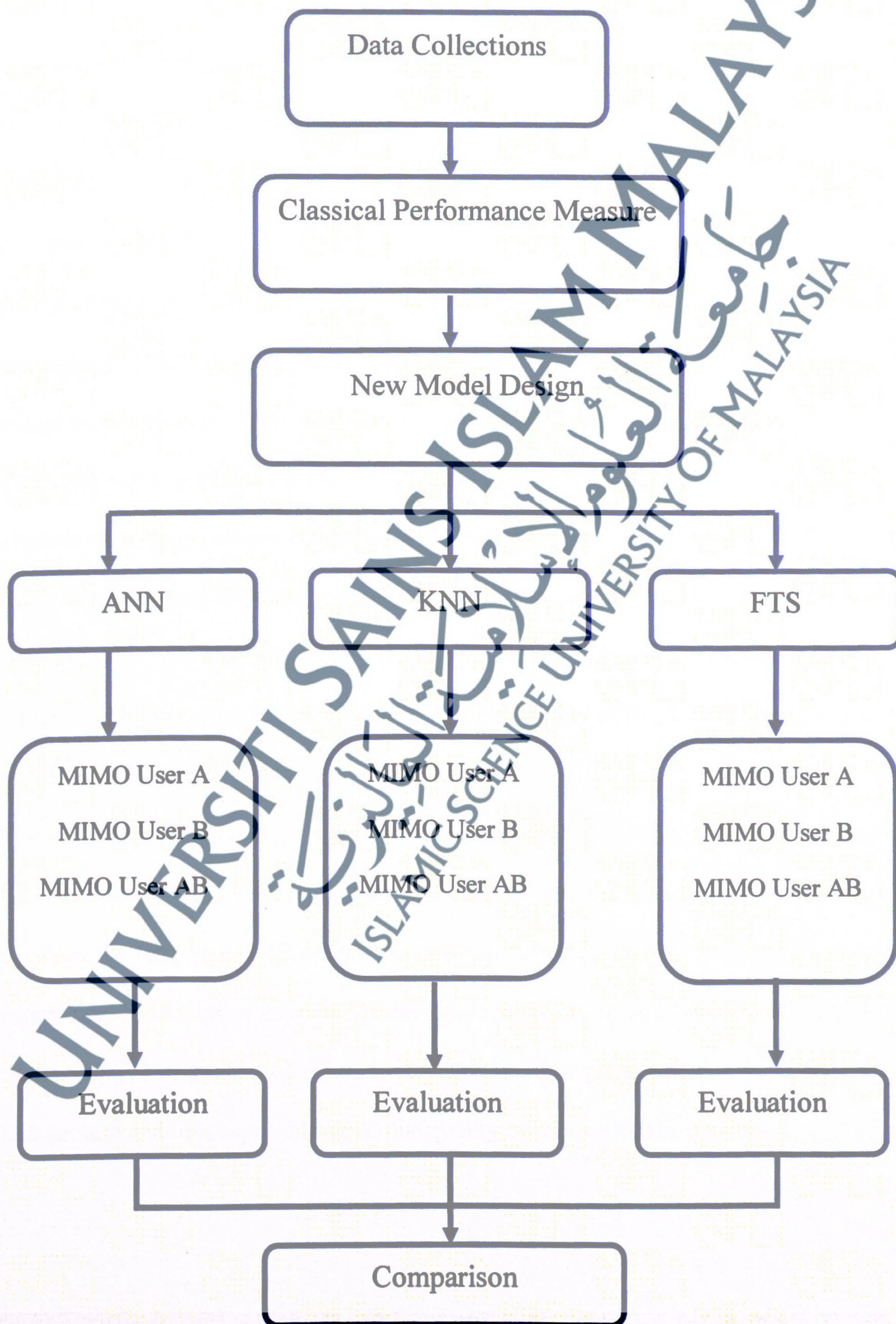
3.0 Introduction

The research methodology is adopted from the systems development life cycle (SDLC) methodology (Ragunath et al., 2010; Basil, 2012; Lekh & Pooja, 2015) that bears a certain resemblance to a methodology rooted in AI research (Cohen & Howe, 1989). Although the problem domain of the study revolves around wireless networking, the solution domain is significantly related to AI. This motivates the usage of the methodology in question.

Essentially, the methodology comprises of five main stages. They are planning, analysis, design, implementation and evaluation. Each stage contributes to the working of the research as a whole. It controls the direction of the study by maintaining a cohesive stance to the established objectives, which is vital in solving the problem at hand.

3.1 Proposed Framework

Figure 16: Proposed Framework



3.1 .1 Data Collection

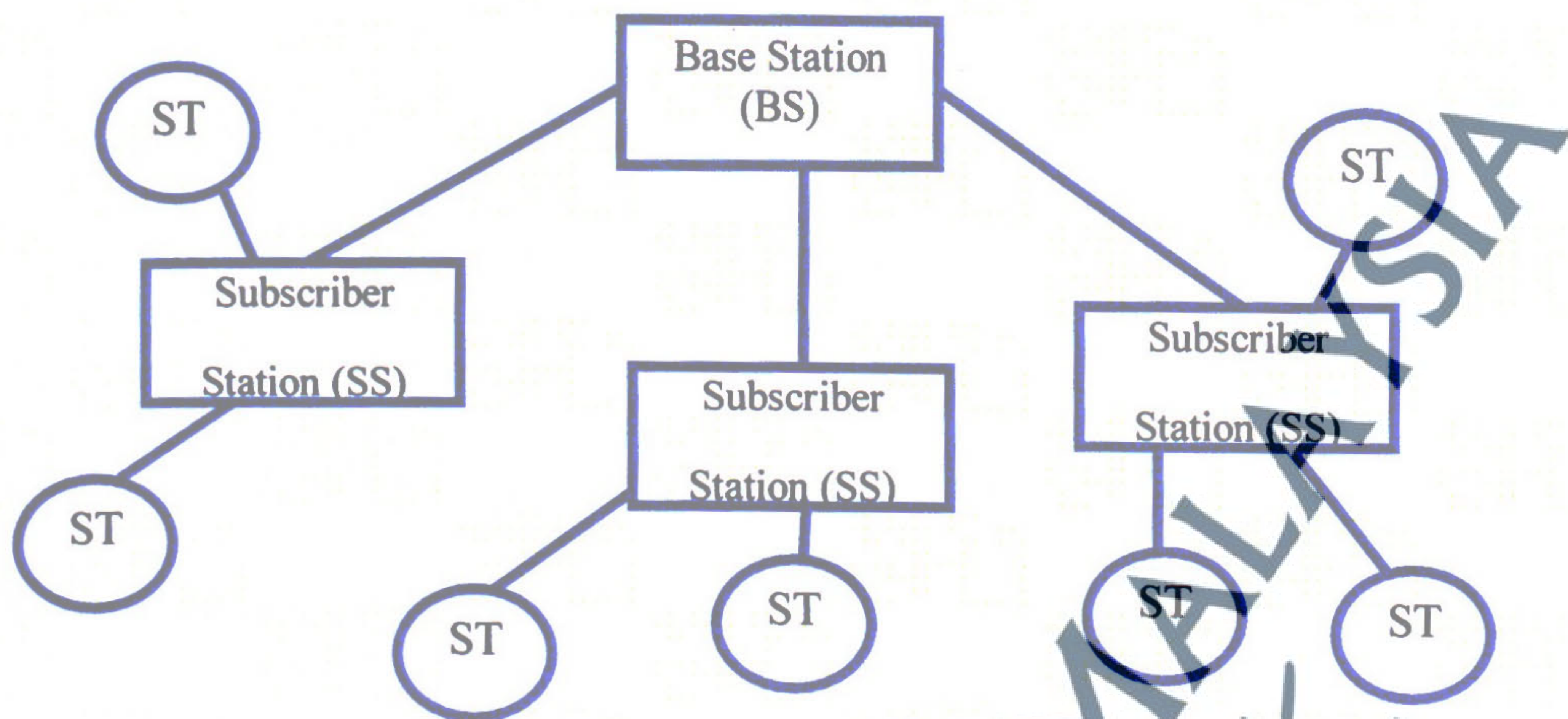
To ensure that the research reflects the actual traffic of a particular WiMAX network (Nandhini & Devarajan, 2014), the actual data is harnessed (Qiang et al., 2007). It is gathered from the traffic of LibyaMax, a WiMAX network that is operated by Libya Telecom and Technology. In other words, the data collection is performed with the help of the company, one of the main telecommunication companies in Libya (Abdulla, 2010).

The inability of the researcher to participate in data collection is non-voluntary. In fact, it is imposed solely by the company. To make it more transparent, it is part of the company policy to prohibit researcher's outsiders from being present within the vicinity of the core technology. This is a strict form of countermeasure to safeguard the trade secret of the company (Villasenor, 2015) from leaking out, intentionally or unintentionally.

The company capitalizes on infrastructures provided by two vendors Huawei and ZTE. Huawei offer their services through the M2000 system while ZTE employ NETNUMEN as the main system for their product. Technically speaking, the data collection is done automatically via the proprietary software provided by the vendors whereby no humanistic intervention is required. Thus, the plausibility of the data is completely reliant on the technology.

The most cost-effective topology is the mesh topology (Carvalho et al., 2013). For this however (Figure 17) there exists a mediating station called a subscriber station (SS) between the base station (BS) and subscriber terminal (ST). The SS plays the role of a router that allows better coverage to the entire network. In addition, the SS can also be made adaptive by generating routes on demand to allow greater flexibility. In the end, the architecture and topology will impact the quality of service for the network.

Figure 17: Mesh Topology (Carvalho et al., 2013).



The real data was collected in a Key Performance Indicator (KPI) is a measurable value that demonstrates how effectively a company is achieving key business objectives. Organizations use KPIs to evaluate their success at reaching targets.

Start time means the date of collecting the data, the period and name of station and using maximum and minimum of online users. Traffic unites by byte, the period time of record the data is 30 minutes.

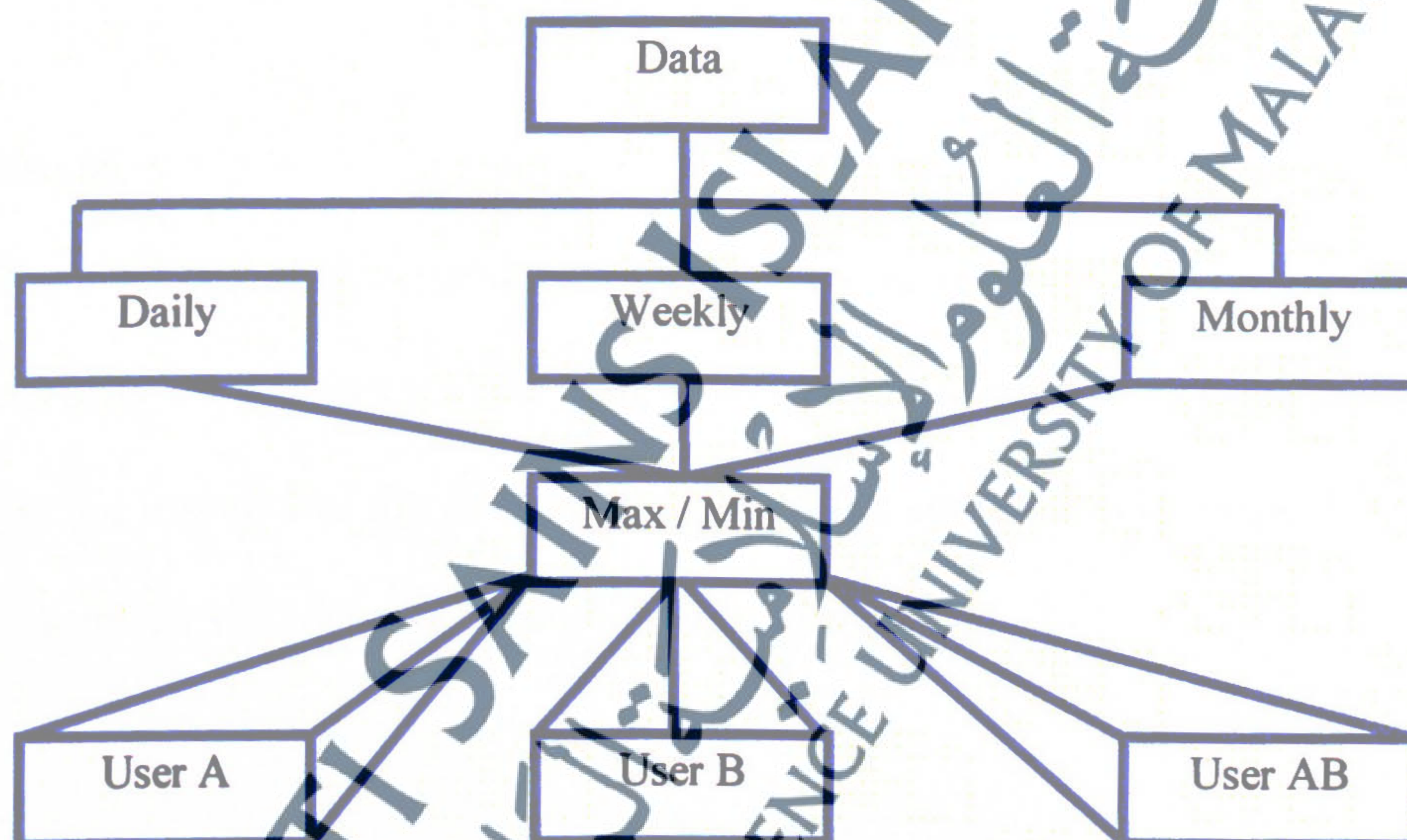
The entire duration of data gathering spans a total of six months or 180 days. During such time, data is collected in its raw form and requires preprocessing to be used in the research. It basically portrays the WiMAX traffic over a certain period of time. For instance, collection can be done in terms of the number of users within a day, week or month.

Data collection is organized in three major temporal categories. They are based on the daily (Lenskiy & Seol, 2012), weekly (Liu & Brown, 2003; Mehrara & Oryoie, 2011), and monthly basis (Joon-Myung et al., 2011; Atsawathawichok et al., 2014). Each

category would also include the maximum (Jing-Ran et al., 2013) and minimum number (Stefanatos et al., 2009) of online users.

Analysis is often conducted based on certain durations. Therefore, data would normally focus on the stipulated category before delving into a certain class of user (Figure 18). To concretize the notion, the research begins by emphasizing on a particular timeline, whether daily, weekly or monthly. For each timeline, the data for the users is enumerated. Thus, they are arranged as daily user A, daily user B, and daily user AB.

Figure 18: Data Taxonomy (Stefanatos et al., 2009)



3.1.2 Classical Performance Measure

The performance of WiMAX traffic forecasting is measured based on the accuracy of prediction. It is actually the difference between the predicted and actual data. To illustrate, a graph is derived from the actual data and used as the benchmark for the prediction. A part of the actual data is then analyzed by the AI approach to make the

prediction. Both are then compared to measure the prediction error, or more specifically, the mean square error (MSE)

3.1.3 Enhanced Model Design (ANN)

1. Planning

Planning entails the breakdown of the research into actionable steps. In other words, this particular stage involves the creation of a plan that allows the research objectives to be fulfilled without any disruption. This is important to ensure the completion of the research within the stipulated timeline.

2. Analysis

The analysis stage examines the artificial neural network (ANN) with greater intensity. It focuses on the viability of employing the method for forecasting. During inspection, the present capacity of the approach with respect to forecasting is noted as well. This is crucial for improvement.

3. Design

Once the current progress on artificial neural network (ANN) is properly established, further enhancement is concocted within the design stage. A new flow for ANN is designed here. It exploits the minimum and maximum value of the test data to make it more conducive to the WiMAX traffic forecasting.

4. Implementation

The implementation involves the development of three approaches in total. They are artificial neural network (ANN), Besides ANN, the two other approaches are

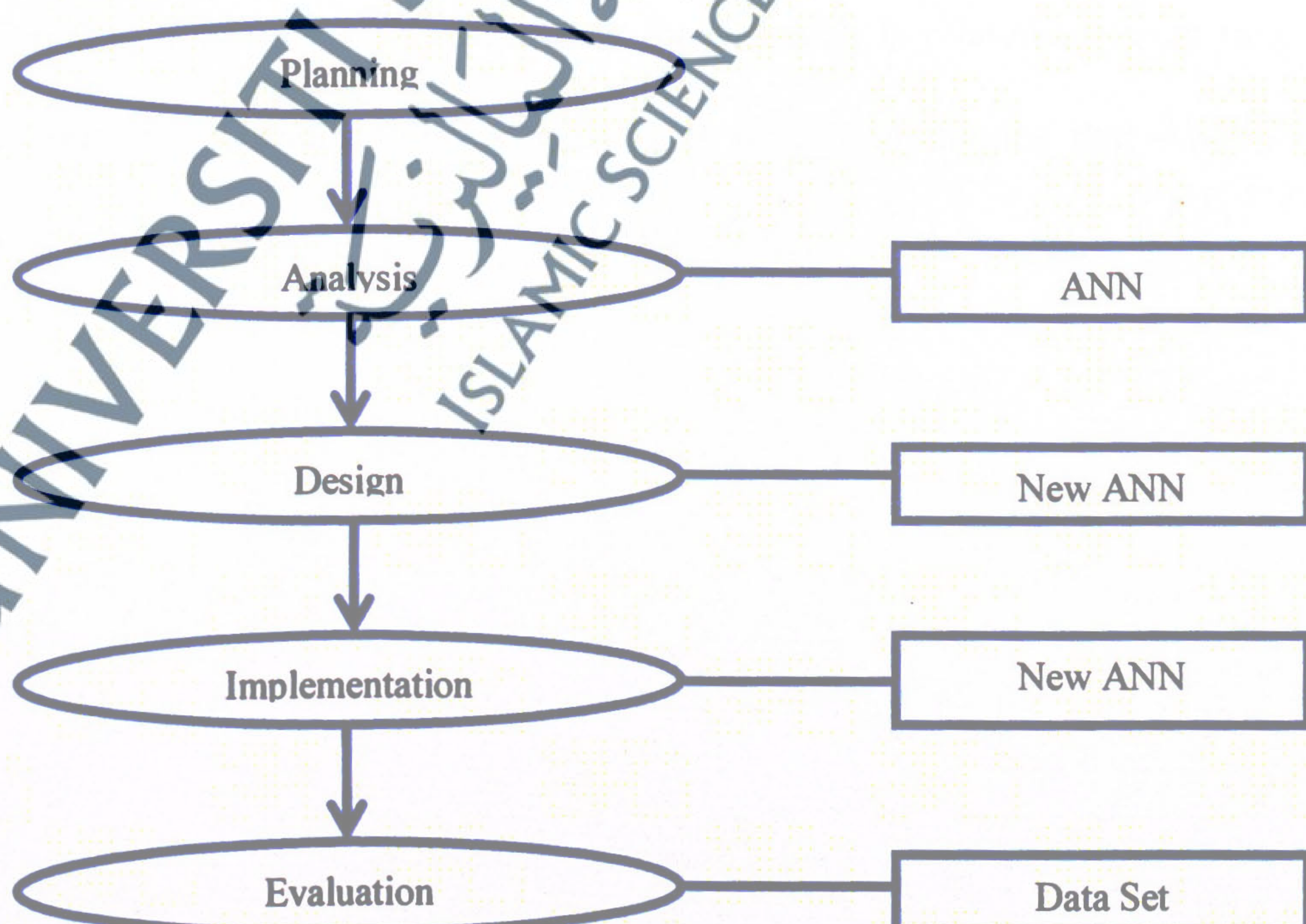
developed to allow a proper comparison to be attempted between them using MATLAB software.

5. Evaluation

In the evaluation stage, all the three approaches are given the same data set and tested in terms of their accuracy on prediction. A standardized measure is vital here. The mean square error (MSE) is a measure that signifies the degree of error exhibited by a particular method. Here, a lower value implies greater accuracy.

By adhering to the methodology (Figure 19) throughout the study, the validity of the research can be maintained more thoroughly. Further elaboration on the research methodology is covered in the following sections. The process flow of each AI approach will be explained in more detail to assert a clearer picture on how their mechanism can vary. However, greater accentuation is given to ANN as compared to the other two.

Figure 19: Research Methodology



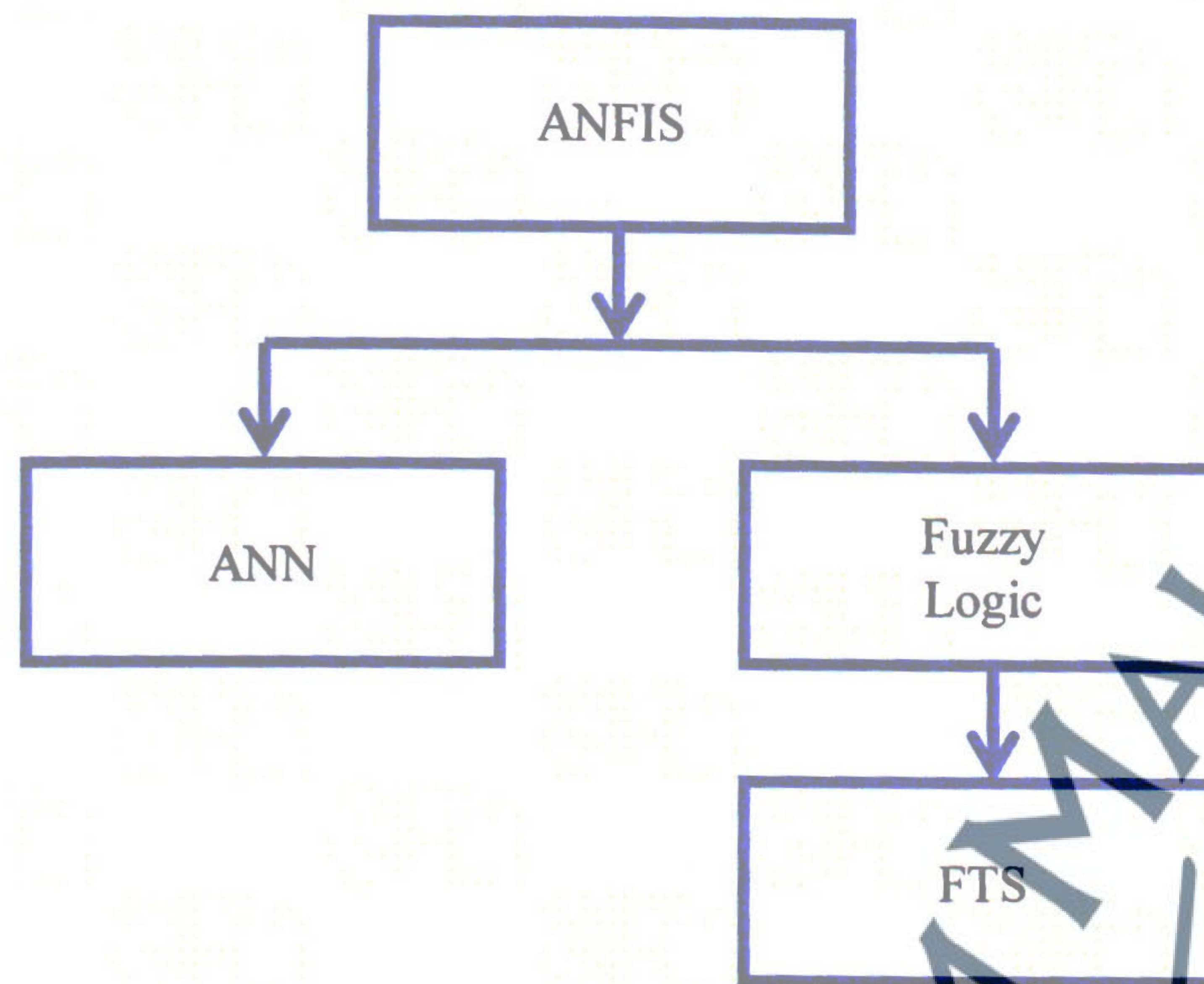
3.1.4 Analysis of Limitation

AI forecasting technique such as ANFIS is superior when compared to its statistical counterpart ARIMA. The difference of performance however is not that significant. In fact, ANFIS demonstrates a marginal enhancement to ARIMA in terms of accuracy. Notice the error of forecasting (RMSE) exhibited by ARIMA and ANFIS. The latter only manages to reduce the extent of error by a small amount as compared to the former.

The computational complexity between the two is rather striking. ANFIS consumes resource at a considerable rate when contrasted against ARIMA. Observe the number of parameters required by ARIMA and ANFIS. The former requires 2, 2 and 7 parameters for all the three cases whereby the latter demands 104, 496 and 1064 parameters. This causes an alarming realization. AI forecasting technique must be made less taxing and more accurate.

To improve the current state of AI traffic forecasting, a series of simplification is needed. Further analysis of ANFIS reveals that it combines two major AI approaches. They are artificial neural network (ANN) and fuzzy logic, which is related to fuzzy time Series (FTS) in traffic forecasting. Theoretically, this revelation implies that ANFIS can be decomposed into two simpler approaches (Figure 20).

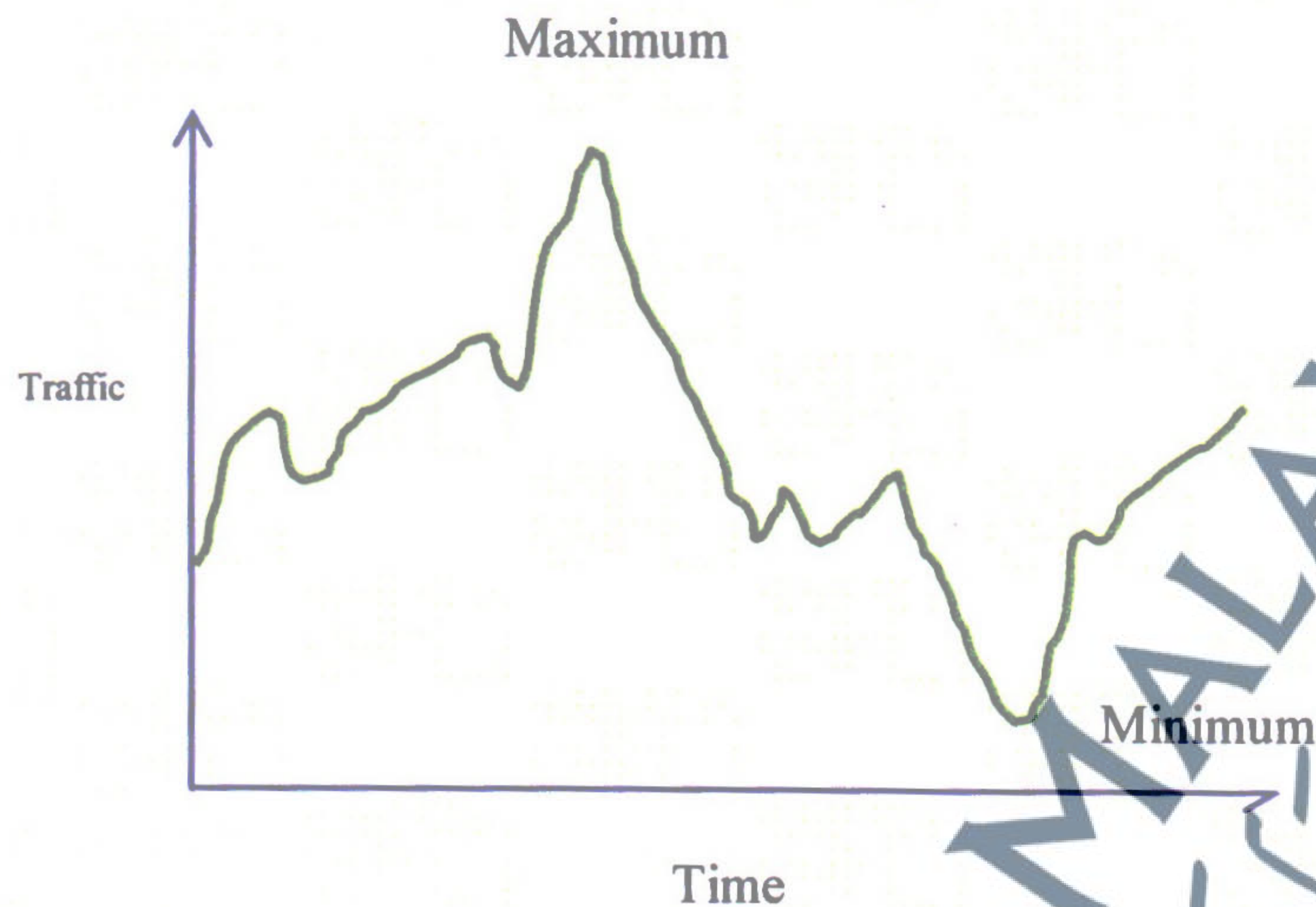
Figure 20: Decomposition of ANFIS



Before progressing further, it must be reminded that the limitation of AI forecasting can be seen from its innate complexity and comparatively scarce predictive improvement. Therefore, the aforementioned analysis suggests the reduction of a complex approach into simpler ones. Simplification reduces the computational complexity of an approach. Rationally, this means that the usage of neural network instead of ANFIS should demand lesser resources.

The simplification however, comes with a price. In traffic forecasting, artificial neural network (ANN) faces a notable difficulty in predicting the peak and nadir of traffic with respect to time (Hall & Mars, 1998; Hall & Mars, 2000). This particular limitation (Figure 21) is attributed to its mechanism. ANN is particularly adept at learning the recurring patterns of traffic. However, the maximum and minimum are usually unique and non-repetitive.

Figure 21: Limitation of ANN in Predicting the Maximum and Minimum



As a way of improving the possible weakness of artificial neural network (ANN) in WiMAX traffic forecasting, the inclusion of other approaches seems necessary. For instance, ANN can be improved by exploiting wavelet transform (Railean et al., 2010) and genetic algorithm. Unfortunately, the employment of these additional methods would inevitably increase the computational complexity of ANN. Therefore, the same dilemma rises yet again.

3.2 Summary

In this chapter, we have explained the steps of the methods research methodology employed in Artificial Intelligent techniques, all of these methods have been introduced sequentially, and the experiments are carried out of each algorithms against the data set defined and design, and evaluate the MSE. These will be details in the next chapter.